

Modeling Population Cell Size Variability in *E. coli* *

Manasi S. Gangan and Chaitanya A. Athale,
Div. of Biology, IISER Pune,
Dr. Homi Bhabha Road, Pashan, Pune 411008, India.
Email: cathale@iiserpune.ac.in

January 15, 2016

Abstract

The sizes in a population of genetically identical *Escherichia coli* cells is known to vary, independent of growth stage. Multiple genes and proteins involved in cell division and DNA segregation are involved in regulating cell size. At the same time, physical factors such as temperature, growth rate and population size also appear to affect cell sizes. How these physical factors interact with the genetically encoded ones to produce such effects is however not clearly understood. Here, we have developed a multi-scale model of bacterial DNA replication coupled to cell division in the context of a logistically growing population. DNA replication is modeled as the stochastic dynamics of replication forks (RFs) which transition probabilistically between two states, stalled and recovered. In the model stalled RFs or incomplete DNA replication results in aberrant cell division. Simulating this model, we demonstrate that the cell-size variability of cultures depends strongly on population size and the growth phase. Our model also predicts the variability in cell sizes is independent of growth temperature between 22° and 42°C. To test the model, we perform experiments

*The work was funded by IISER core-funding and a Basic Biology Grant by the Dept. of Biotechnology, Govt. of India BT/PR1595/BRB/10/1043/2012. MSG was funded by a fellowship by the Indian Medical Council for Research (ICMR).

with *E. coli* strains mutant for *recA*, *sulA* and *slmA*, factors known to affect replication fork dynamics and coupling to cell division. Our validated model of RF stalling based cell division, can reproduce the variability in cell size distribution seen in these mutant strains, and provides a predictive tool to further examine stochastic effects in bacterial cell size regulation.

Keywords: cell size; bacterial cell cycle model; population variability; stochastic replication

1 Introduction

Size and shape cells are characteristic for a given cell type and is a complex phenotype, depending on multiple genetic and non-genetic factors. Understanding the non-genetic variability of cell size in a population, could provide an indication of the mechanisms regulating it. However to do this, it is essential for a quantitative theoretical framework to be developed. Cells of the model bacterium *Escherichia coli* have been described as spherocylinders with a length of $2\text{ }\mu\text{m}$ and width of $1\text{ }\mu\text{m}$ at birth. A single newborn *E. coli* cell during growth elongates to twice its length and divides symmetrically into two daughter cells. Cells sampled at the same growth stage (e.g. birth) are variable primarily in cell length (L), while the width appears to be constant as seen in measurements made in light- and electron-microscopy [1, 2, 3]. The frequency distribution of cell lengths typically shows a positive skew due to the presence of long cells which have been defined in literature as having $L > 8\text{ }\mu\text{m}$. Environmental factors such as low bacterial density [4] or a shift to richer media [5] have been shown to increase the proportion of long-cells. The average cell mass and DNA content is independent of temperature (25°C and 37°C), but a change in the growth medium can result in change in mass and DNA content as seen in *Salmonella typhimurium* [6] and *E. coli* [7]. On the other hand, a study on the effect of temperature on *E. coli* sizes showed cells at 22°C are shorter than at 37°C [8]. At the same time, growth rate alone has been shown to correlate with increased cell size and multiple nucleoids in *Salmonella* [6]. These observations with respect to the effect of temperature and nutrients on cell size imply gaps in both measurement and our theoretical understanding. Recent improvements in light microscopy now allow visualizing sub-cellular processes and their dynamics even in cells as small as *E. coli* [9]. Coupled with computational image-analysis methods

[10], a systematic quantification could result in a disambiguation of these results.

The molecular mechanisms by which nutrient availability determines bacterial cell size has recently been shown to be mediated in *E. coli* by phosphoglucose mutase (pgm) [11] and glucosyltransferase OpgH [12] and in *Bacillus subtilis* by the protein UgtP [13]. While these factors are known to affect the central tendency of cell size, they do not explain effects that change the spread of the distributions. Additionally most pathways regulating *E. coli* cell size, including nutrient sensors, converge on FtsZ as the primary molecular determinant of newborn cell size [14, 15, 16]. The regulation of division in *E. coli* by oscillations of MinCDE proteins [17, 18] and nucleoid occlusion by the SlmA-DNA complex [19, 20] is well established. Additionally an indirect effect has been observed based on Sula inhibition of FtsZ, triggered by RecA dependent SOS response [21]. While elongated cells (length $> 8 \mu m$) in wild-type populations have been reported, recent quantification of the cell-division septum has estimated a deviation from the mid-cell by 10% of cell length [22]. Theoretical modeling could serve to reconcile the relative importance of these mechanisms and lead to a deeper understanding of the regulation of a complex phenotype such as cell size.

Mathematical models of bacterial cell length distributions in growing populations based on single cell growth rates were proposed by Collins and Richmond to result in stable cell length distributions in *Bacillus cereus* [23]. Variability in cell lengths was modeled by assuming either asymmetric cell division [1] or stochastic partitioning of molecular components during cell division [24]. Recent developments in microfluidics have allowed the measurement of single-cell bacterial growth kinetics [25], making it possible to distinguish between the “timer” and “sizer” mechanisms that had been previously proposed to govern bacterial cell sizes [26]. Recently, an analytical model which invoked volume change as a governing criterion for cell size determination, and referred to as an “incremental model”, has been consistent with all the data available on cell size and its dynamics [27, 28]. This model is however phenomenological, lacking in a molecular mechanism for the source of stochasticity in a population of genetically identical cells.

Here, we have developed a model of *E. coli* cell size in a growing population of simulated cells. The model spans the scales of sub-cellular replication fork dynamics, whole cell scale birth, growth and division dynamics and population growth of cells in a limited nutrient pool. Experimental measurements of cell size distributions of *E. coli* mutants of the SOS response

pathway, are used to test the model predictions.

2 Theory: Multi-scale model of bacterial population growth

A multi-scale model of a population of individual cells or ‘agents’ growing in length and dividing has been developed. The instantaneous growth rate (doubling time) is modeled as a variable, based on the logistic model, which depends on the size of the population scaled by the available nutrients. This reproduces the growth phases of a population growing in the presence of limited nutrients seen in typical batch cultures- lag, log and stationary. A modified birth, chromosome-replication and division (BCD) model of the bacterial cell cycle has been developed, in which chromosome-replication has been explicitly modeled by the discrete, stochastic dynamics of replication forks (RFs). RF dynamics is coupled to cell division through a model of cell division inhibition, based on the role of RecA-SulA inhibition of FtsZ. Model parameters are taken from literature where available, and the remaining are optimized (Table 1).

2.1 Stochastic multi-fork replication

DNA replication in *E. coli* is modeled by a two-state dynamics of an individual RF. An RF is either in a stalled or replicating state (Figure 1(a)). In the replicating state, DNA is replicated by the RF with a genome replication speed (v_r). The value of v_r is taken from the *E. coli* DNA PolIII holoenzyme replication speed [29, 30]. The frequency of stalling (f_s) determines the transition of an RF from replicating to stalled state and the frequency of recovery (f_r) determines the transition back from a stalled to a replicating state. The values of f_s and f_r were optimized to qualitatively reproduce experimental cell length ranges in wild-type cells. These frequencies are compared at each iteration step to a random number drawn from $U[0, 1]$ to determine the state of the RF.

During rapid growth, *E. coli* is known to initiate more than one pair of RFs (Figure 1(b)). The number of RFs (n_{RF}) at any given time in a cell is modeled by modifying Cooper and Helmstetter’s model of multi-fork replication [31] as:

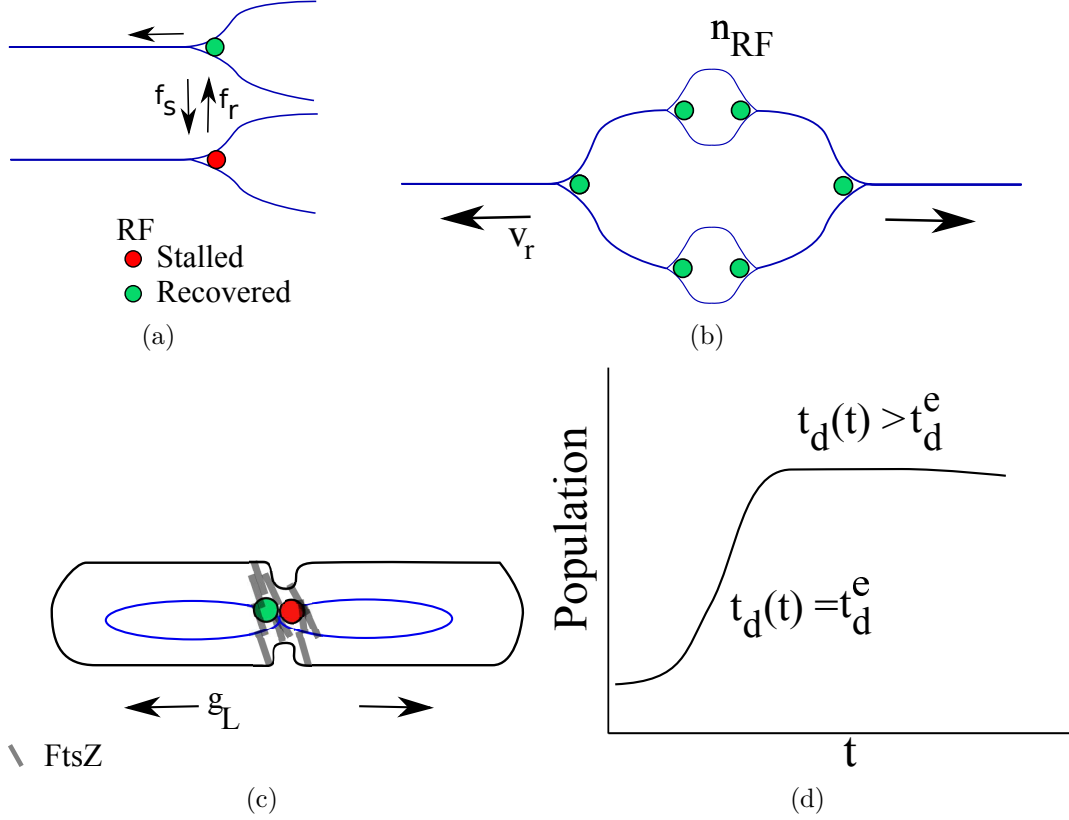


Figure 1: (a) DNA replication can proceed if the replication fork (RF) is in the recovered state (green) but not if it is in the stalled (red) state. The transition frequency of forks to a stalled state is determined by f_s , the frequency of stalling, and the transition to recovered state by f_r , the frequency of recovery. (b) Replication proceeds at a speed given by v_r bidirectionally on the genome. (c) Cell growth is governed by the growth rate (g_L) and ftsZ assembly leads to cell division. The cell can only divide if all RFs are in recovered (green) state. (d) The population size (N) as a function of time (t) follows the logistic model due to a variable doubling time ($t_d(t)$) based on Equation 4 in this schematic plot. The growth curve enters the ‘stationary phase’ when $t_d(t)$ exceeds t_d^e due to the population size ($N(t)$).

$$n_{RF} = \begin{cases} 2 & \text{if } t_d(t) \geq C + D \\ 2 \cdot (1 + (C + D)/t_d(t)) & \text{if } t_d(t) < C + D \end{cases} \quad (1)$$

where C and D is the time taken for chromosome replication and cell division respectively and $t_d(t)$ is the instantaneous doubling time. The C- and D-periods are assumed to be uniform based on previous measurements in fast growing cultures [32].

2.2 Cell elongation model

The growth rate of a single cell ($g_L(t)$) is modeled to depend on the instantaneous doubling time ($t_d(t)$) and the birth length (L_b) of the cell as:

$$g_L(t) = L_b/t_d(t) \quad (2)$$

This constant growth rate results in a single linear rate of cell length increase (Figure 1(c)). This linear assumption averages the recently demonstrated bi- and tri-linear fits from single cell growth experiments [33]. Here we model cell elongation rate to change with the growth phase of the population. At time t , with a given doubling time ($t_d(t)$), the value of $g_L(t)$ is constant for all cells in the population, consistent with the “incremental model” of bacterial cell growth [27].

2.3 Cell division model

The simulated cell can divide ($\Gamma = 1$) depending on the cell age (t_{cell}), the length of the genomic DNA in the cell (DNA_{cell}) and the state of the n^{th} replication fork ($RF_n = 0$ or 1), as given by:

$$\Gamma = \begin{cases} 1 & \text{if } \begin{cases} t_{cell}(t) \geq t_d(t) \\ DNA_{cell}(t) = 2 \cdot DNA_{Ecoli} \\ RF_{n_{RF}} = 1 \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $t_d(t)$ is the the instantaneous doubling time of cells, DNA_{Ecoli} is the complete genome length of *E. coli* and n_{RF} refers to the identity of the replication forks (2 to n_{RF}). The cell age criterion accounts for the BCD cell cycle, the genome elongation for nucleoid occlusion and the RF state models the inhibition of division by the SOS-pathway. In the model, when $\Gamma = 1$ cell division proceeds instantaneously, since we assume rapid kinetics of FtsZ polymerization (Figure 1(c)). In this way, our cell division model couples the stochastic DNA replication with the deterministic parts of the model. In the

model, cell length variation beyond the expected range (2 to 4 μm) arises from a failure of cells to divide. Such an elongated cell may divide at the next round of replication if the conditions are satisfied. The model does not impose a ‘sizer’ or ‘timer’ mechanism, instead a cell size ‘emerges’ from the cell cycle model.

2.4 Population dynamics model

The growth of a population of discrete cells with finite resource is modeled by a variable doubling time ($t_d(t)$) (Figure 1(d)). Based on the logistic growth model, the doubling time for the whole population at each time point t is calculated as:

$$t_d(t) = \frac{t_d^e}{(1 - N(t)/K)} \quad (4)$$

The value of the exponential doubling time (t_d^e) in the model is constant for a specific environmental condition- temperature or nutrient medium. We have experimentally measured t_d^e of *E. coli* in LB grown at 22°, 30°, 37° and 42°C (Table 2). The carrying capacity (K) is fixed for all calculations (Table 1). As a result the population undergoes the expected logistic growth (Figure 2(a)).

3 Materials and methods

3.1 Bacterial growth

Single colonies of *E. coli* MG1655 cells (CGSC, Yale) were picked and grown overnight in LB medium (Difco Labs, Mumbai, India) at 37°C with constant shaking at 170 rpm (Forma, ThermoScientific, USA) and inoculated at 1% in a 100 ml culture for growth at 22°, 30°, 37° and 42°C. Growth curves were measured by monitoring O.D. at 600 nm (Biophotometer, Eppendorf, Germany) and the \log_{10} (cell density) as a function of time was determined (1 O.D. 600 nm = $8 \cdot 10^8$ cells/ml). The growth rate (r) and doubling times (t_d) were determined for each temperature (Table 2). from the \log_{10} (cell density) plot with time. The slope of the linear region between time points t_1 and t_2 was used to determine the growth rate (r) in *cells/ml · min* by $r = [\log_{10} N(t_1) - \log_{10} N(t_2)]/[\Delta t \cdot \log_{10} 2]$, where $\Delta t = t_2 - t_1$. The exponential phase doubling time is related to growth rate by $t_d^e = 1/r$ [34].

3.2 Microscopy and image analysis

Growth experiments were performed with *E. coli* MG1655, $\Delta recA$, $\Delta slmA$ and $\Delta sulA$ strains (CGSC, Yale) in LB at 37°C as described in Section 3.1. Cell sizes were measured as described previously [35] and is briefly summarized here. Cells from the growth experiment were sampled at multiple time-points, fixed in paraformaldehyde, mounted on glass-slides with coverslips, imaged in differential interference contrast microscopy using 40x, Zeiss AxioImager Z1 (Carl Zeiss, Germany) and cell lengths measured using a routine developed in-house in MATLAB (Mathworks Inc., Natick, MA, USA). Cell length variability was estimated using the fano factor (FF) and coefficient of variation (CV) of cell lengths, where $FF = \sigma^2/\mu$, and $CV = \sigma/\mu$, where σ^2 is the variance, σ is the standard deviation and μ the mean of the distribution. Percentage long cells are estimated by finding the proportion of cells exceeding 8 μm length.

3.3 Simulations

The simulation was written in MATLAB (Mathworks Inc., Natick, MA, USA) and run on a machine with two 2.93 GHz Quad-Code Intel Xeon processors with 16GB RAM. A single run with $N(t = 0) = 10$ and $K = 1000$ cells respectively and time step $\delta t = 1$ second required 6 hrs to run on an average.

4 Results

4.1 Population dynamics and multi-fork replication

We initially simulated the effect of temperature on the cell size variability in a growing population, by using experimentally measured values of t_d^e to modulate population growth kinetics. The populations with the smallest doubling time (42°C, $t_d^e = 18$ min) grew the fastest, followed by slower growth rates for lower temperatures (Figure 2(a)). The discrete model shows occasional jumps in cell numbers, due to the discrete nature of the population model, but follows the shape of the logistic curve. While only the fastest growing population at 42°C reaches saturation in the 4 hours of simulation, our results depend primarily on the log phase of growth. The single cell elongation rate (g_L) reduces linearly with increasing population size (Figure 2(b)), as

Table 1: Simulation parameters

Symbol	Parameter description	Value	Reference
K	Carrying capacity	10^3 cells	this study
N_i	Initial population size	10 cells	this study
t_d^e	Maximal doubling time	18, 25, 30, 60 min	Table 2
f_s	Frequency of RF stalling	0.005 s^{-1}	this study
f_r	Frequency of RF recovery	0.09 s^{-1}	this study
L_g	Genome length	$4 \cdot 10^6$ bps	[36]
v_r	Genome replication speed	10^3 bps/s	[29, 30]

expected from the $g_L \propto 1/t_d$ dependence (Equation 2), and in turn the dependence of $t_d \propto 1/N(t)$ (Equation 4). In the representative snapshots from growth at 37°C ($t_d^e = 25$ min) taken every 0.8 hours, in the early phase of rapid growth, each cell has 8 RFs (upto 1.6 hours) (Figure 2(c)). As the population growth slows due to saturation effects, a shift in the distribution of RFs per cell ($n_{RF} = 6$ to 8) is seen in 2.4 to 3 h, and near at 4 hours, the majority of cells have $n_{RF} = 2$ when growth has reached saturation. Cell sizes of such a population at the start of the simulation are initialized to $2\text{ }\mu\text{m}$. As the population grows, the cell lengths begin to be more broadly distributed (Figure 2(d)). The cell length distributions were fit by a lognormal function, in which the variance increases with time and appears to saturate. The lognormal variance is maximal at 3.2 hours ($\sigma^2 = 2.37\mu\text{m}^2$) and minimal at 0.8 hours from the snapshots. The multi-scale model thus reproduces the expected population behavior of cell numbers with time at different temperatures. The observed changes in replication fork and cell length distributions suggest a growth phase dependence, which we have analyzed through the growth of the population.

4.2 Temperature and growth phase dependence of cell size variability

The simulated population cell length distributions appear most variable in the mid-log phase, as seen from the spread in the length frequency distribution snapshots (Figure 2(d)). On estimating the percentage of elongated cells (cells with length $> 8\text{ }\mu\text{m}$) throughout the entire time-course of the simulation, we find an initial increase, which saturates as a function of time

Table 2: Experimental measures of doubling time as a function of temperature at which *E. coli* was grown.

Temperature	Doubling time (min)
22°C	65.3 ± 0.8
30°C	29.2 ± 0.4
37°C	25.7 ± 0.1
42°C	17.9 ± 0.2

(Figure 3(a)). The mean %age long cells of the 42°C scenario reaches saturation earlier than 37°C and followed by 30°C. The %age long cells of the slowest growing population at 22°C, does not saturate at all. When the data was plotted as a function of increasing population size, the plots from multiple temperatures appear more comparable to each other, as compared to the time-dependent plots. Similarly, estimating cell length variability by the fano factor, FF (Figure 3(b)) and the coefficient of variation, CV (Figure 3(c)) as a function of time, shows an increasing trend. The time-dependent graphs of both measures are offset, such that the CV and FF are lower, with decreasing temperature. Plotting FF and CV as a function population size however collapses all the curves onto the same profile. This suggests the cell length variability in simulations for the different temperatures is dominated by the population size. The difference seen in the time dependent profiles can be understood in terms of the logistic growth model where temperature determines the maximal doubling time (t_d^e), while instantaneous population size ($N(t)$) determines the corresponding instantaneous doubling time ($t_d(t)$), which in our model determines n_{RF} . As a result, cell length variability, when scaled by population size, follows the same trend across the temperatures. Hence in our model of growth rate and replication stochasticity driven cell size variability, cell size variability is similar for all the growth temperatures modeled (22°C to 42°C), but depends on population size. We proceed to test our proposed model mechanism by perturbation of replication dynamics and compare the outcome with experimental measurements.

4.3 Replication fork stochasticity and cell size variability

In the two-state model of RFs, the stalling frequency (f_s) and recovery frequency (f_r) determine the average state of each RF and result in a chromosomal replication time. The RF stalling frequency was assumed to be 0.005 s^{-1} , almost an order of magnitude smaller than the *in vitro* measurement of *E. coli* Pol-III holoenzyme $f_s = 0.02\text{ s}^{-1}$ [37]. The value of f_s was optimized for qualitative agreement of simulated cell lengths with the experimental measures from *E. coli* wild-type cells grown at 37°C in LB. To simulate the effect of a mutant with reduced ability of stalled RFs to inhibit cell division, we reduced f_s to 0.001 s^{-1} , based on the assumption such cells sense RF stalling at a lower frequency as compared to wild-type (Figure 4(a)). As expected, the simulated cell length variability (FF) is lower than the ‘wild-type’ throughout the growth period (Figure 4(a)). To simulate a genetic mutation in the SOS response pathway such as a *recA* deletion, we decreased the frequency of replication fork recovery (f_r) by two-fold in simulation, since in the absence of *recA* we expect replication fork recovery to be less frequent. The cell length variability measures expectedly increased through all stages of the growth curve. The FF as a function of time and population size (N) from these simulations showed the same trend, with ‘mutant’ scenarios appearing distinct from the ‘wild type’ (Figure 4(b)). To test these model predictions, growth experiments were performed and microscopic images of cells in wild-type (*E. coli* MG1655), $\Delta\textit{recA}$ (lower f_r), $\Delta\textit{sulA}$ (lower f_s) and $\Delta\textit{slmA}$ (lower f_s) were acquired (Figure 4(c)). Using our previously developed routine for cell length analysis, the FF of population cell length distributions as a function of cell density were estimated (Figure 4(d)). The trends of $\Delta\textit{recA}$ strains showing higher variability and $\Delta\textit{sulA}$ and $\Delta\textit{slmA}$ showing lower variability, qualitatively match those predicted by simulation. Thus our model predictions are validated of a RF stalling frequency and cell size variability model. It suggests the coupling of multi-fork replication and cell division, can drive cell size variability in genetically identical cells of *E. coli* as a function of cell density.

5 Discussion

We have developed a novel model of bacterial cell growth and division, which integrates sub-cellular processes with population dynamics. The agent-based multi-scale model integrates a stochastic, multi-fork chromosome replication model coupled to cell division, in the context of a growing population with a finite carrying capacity. Using this model, we examine the effect of growth phase and temperature on cell size variability, to disambiguate previous experimental findings. The only source of stochasticity in the model is the probabilistic transition of DNA RFs between stalled and recovered states. Cells become filamentous, i.e. long, when they either do not complete DNA replication within a bacterial cell cycle or contain one or more stalled RFs at the time of division. The cell length growth rate ($g_L(t)$) on the other is assumed to be uniform for all cells in the population for a specific time point (t). The model predicts cell length variability is highest in the exponential growth phase as compared to the lag and stationary phases. The time of onset of increase in cell length variability from the lag to log phases differs between cultures simulated at growth temperatures 22° to 42°C. However, when scaled by population size, the variability collapses onto the same curve across temperatures. The sensitivity of the model is tested to changes in the RF frequencies of stalling (f_s) and recovery (f_r). An increased time spent by RFs in the stalled state results in increased variability in cell sizes. A comparison with experimental measurements of growth phase dependent cell size variability of deletion mutants for proteins mediating the DNA RF recovery (SOS response factors) and nucleoid occlusion proteins validates our model.

Due to the small number of parameters (Table 1), this model is relatively tractable. Some parameters such as genome elongation rate, maximal doubling time at different temperatures (Table 2) are derived from experimental measurements in previously published work or measured in this study. Other parameters such as the frequencies of RF stalling and recovery are optimized to fit the extent of cell size distribution observed experimentally in wild-type cells. The f_s of the *E. coli* Pol-III holoenzyme has been estimated using *in vitro* reconstitution to be 0.02 1/s [37]. However, when such a high value was used in our model, it resulted in a far higher frequency of elongated cells, as compared to experiments, and as a result was not used. The lower frequency of stalling *in vivo* as compared to *in vitro* could be the result of the fact that *in vivo*, additional proteins such as RecA and RecBCD [38], helicases, single strand binding proteins and resolvers [39], which are all implicated in RF

recovery pathways. This could also explain why cells of the $\Delta recA$ mutant are only slightly more variable than wild type cells (Figure 4(d)). In future *in vivo* single-molecule measurements that have been developed to follow replication proteins [40], combined with systematic screening of mutants, could be used to quantify RF dynamics more precisely.

The multi-fork replication model implemented here is based on the Cooper and Helmstetter model [31], which predicts a change in n_{RF} with changing growth rate. Our simulations demonstrate the early growth phase of cells contains multiple RFs (up to 8) and saturation phase cultures result in only a pair of RFs per cell. Importantly, the probability of replication stalling is modeled as an independent event, but cell division evaluates the total number of stalling events. As a result, cells experience a multiplicative effect of RF stochasticity on cell division. As a consequence, rapid growth results in higher cell size variability as compared to saturated growth. This clarifies the apparent contradictions in previous reports of temperature [8, 7] and nutrient dependence [41] of cell sizes, where it was unclear which factor primarily determined the distribution. We hypothesize that a simple explanation based on multiplicative probabilistic events, based on multi-fork replication underlie these effects. In future, experiments to count the RF numbers under these conditions could further test our hypothesis.

Single-cell quantitative studies of bacterial growth have begun to reveal subtle effects of asymmetric division and ageing in *E. coli* due to improvements in microscopy [9]. A theoretical model of “increment” based control of cell size of *E. coli* and *Caulobacter crescentus* has been described [27]. The variability of growth rates observed in single-cell continuous culture [25] was explained by variations in the cell elongation rate (g_L) and cell size added [28], further confirming the “increment” model. While this model is consistent with experimental data, it lacks a molecular mechanism, that could cause such variation. Additionally in the absence of collective effects, the behavior only addresses single cell variability. Here in our work we have used a multi-scale approach to address both the morphological measurements seen in older work, as well as propose molecular processes that drive it.

Discrete models of bacterial populations have been widely used in the past to examine spatial patterns [42], chemotaxis [43], synthetic biology [44], branching growth of colonies [45] and population effects in quorum sensing [46]. However this is the first report of a replication-division coupling model that results in cell shape distribution as a function of population size and growth rate.

Our model of replication stochasticity and the effect of multi-fork replication on cell size variability predicts population cell size variability, which is consistent with previous work. Our results also disambiguate some of the results obtained with respect to cell size and temperature, suggesting within a range (22 to 42°C) population cell size variability is more strongly determined by growth rate as compared to temperature. Additionally by performing simulated mutation experiments, we find a qualitative match between simulations and experiments performed on genetic mutants. In conclusion this study could form the basis to use cell size variability to better understand the interactions between regulatory factors governing a phenotype as complex as cell size.

6 Acknowledgements

MSG is supported by a fellowship by the Indian Council of Medical Research. Hemangi Chaudhari was involved in the initial conceptualization of the project. IISER Pune core funding supported the project.

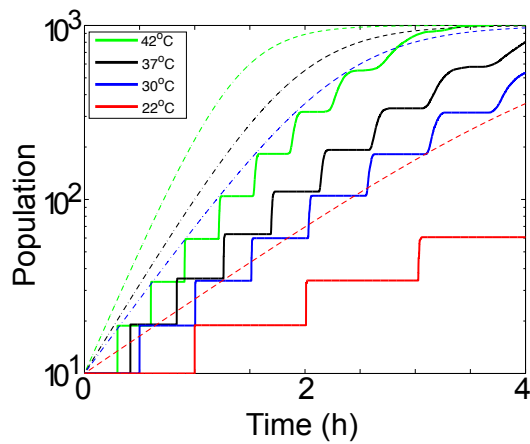
References

- [1] J. Cullum, M. Vicente, J. Bacteriol. **134**(1), 330 (1978)
- [2] N.B. Grover, C. Woldringh, L. Koppes, J. Theor. Biol. **129**(3), 337 (1987)
- [3] T. Kaya, H. Koser, Phys. Rev. Lett. **103**(13), 138103 (2009)
- [4] F. Maclean, R. Munson, J. Gen. Microbiol. **25**, 17 (1961)
- [5] H.E. Kubitschek, J. Bacteriol. **172**, 94 (1990)
- [6] M. Schaechter, O. Maaloe, N.O. Kjeldgaard, J. Gen. Microbiol. **19**(3), 592 (1958)
- [7] T.E. Shehata, A.G. Marr, J. Bacteriol. **124**(2), 857 (1975)
- [8] F.J. Trueba, E.A.V. Spronsen, J. Traas, C.L. Woldringh, Arch. Microbiol. **131**(3), 235 (1982)
- [9] X. Wang, R. Reyes-Lamothe, D.J. Sherratt, Biochem. Soc. Trans. **36**(Pt 4), 749 (2008)
- [10] R. Eils, C. Athale, The Journal of cell biology **161**(3), 477 (2003). DOI 10.1083/jcb.200302097
- [11] N.S. Hill, R. Kadoya, D.K. Chatteraj, P.A. Levin, PLoS Genet. **8**(3), e1002549 (2012)
- [12] N.S. Hill, P.J. Buske, Y. Shi, P.A. Levin, PLoS Genet. **9**(7), e1003663 (2013)
- [13] R.B. Weart, A.H. Lee, A.C. Chien, D.P. Haeusser, N.S. Hill, P.A. Levin, Cell **130**(2), 335 (2007). DOI 10.1016/j.cell.2007.05.043
- [14] J.F. Lutkenhaus, H. Wolf-Watz, W.D. Donachie, J. Bacteriol. **142**(2), 615 (1980)
- [15] E. Bi, J. Lutkenhaus, Journal of Bacteriology **172**(5), 2765 (1990)
- [16] J. Lutkenhaus, Annu. Rev. Biochem. **76**, 539 (2007). DOI 10.1146/annurev.biochem.75.103004.142652

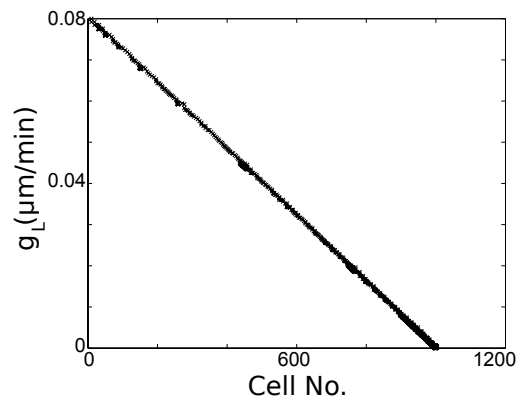
- [17] D.N. Raskin, P.A.J. De Boer, Proc. Nat. Acad. Sci. USA **96**(April), 4971 (1999)
- [18] A. Varma, K.C. Huang, K.D. Young, J. Bacteriol. **190**(6), 2106 (2008). DOI 10.1128/JB.00720-07
- [19] C. Woldringh, E. Mulder, J. Valkenburg, F. Wientjes, A. Zaritsky, N. Nanninaga, Res. Microbiol. **141**(1), 39 (1990)
- [20] T.G. Bernhardt, P.a.J. de Boer, Mol. Cell **18**(5), 555 (2005). DOI 10.1016/j.molcel.2005.04.012
- [21] O. Huisman, R. D’Ari, S. Gottesman, Proc. Natl. Acad. Sci. U.S.A. **81**(14), 4490 (1984)
- [22] J.M. Guberman, A. Fay, J. Dworkin, N.S. Wingreen, Z. Gitai, PLoS Comp. Biol. **4**(11), e1000233 (2008). DOI 10.1371/journal.pcbi.1000233
- [23] J.F. Collins, M.H. Richmond, J. Gen. Microbiol. **28**, 15 (1962)
- [24] D. Huh, J. Paulsson, Nat. Gen. **43**(2), 95 (2011). DOI 10.1038/ng.729
- [25] P. Wang, L. Robert, J. Pelletier, W.L. Dang, F. Taddei, A. Wright, S. Jun, Curr. Biol. **20**(12), 1099 (2010). DOI 10.1016/j.cub.2010.04.045
- [26] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert, M. Doumic, BMC Biol. **12**, 17 (2014). DOI 10.1186/1741-7007-12-17
- [27] A. Amir, Phys. Rev. Lett. **112**(May), 1 (2014). DOI 10.1103/PhysRevLett.112.208102
- [28] S. Taheri-Araghi, S. Bradde, J.T. Sauls, N.S. Hill, P.A. Levin, J. Paulsson, M. Vergassola, S. Jun, Current Biology **25**(3), 385 (2014). DOI 10.1016/j.cub.2014.12.009
- [29] M. O’Donnell, P. Studwell, J. Biol. Chem. **265**(2), 1171 (1990)
- [30] T.M. Pham, K.W. Tan, Y. Sakumura, K. Okumura, H. Maki, M.T. Akiyama, Mol. Microbiol. **90**(2), 584 (2013). DOI 10.1111/mmi.12386
- [31] S. Cooper, C.E. Helmstetter, J. Mol. Biol. **31**(3), 519 (1968)

- [32] O. Michelsen, M.J.T. de Mattos, P.R. Jensen, F.G. Hansen, *Microbiol.* **149**, 1001 (2003). DOI 10.1099/mic.0.26058-0
- [33] G. Reshes, S. Vanounou, I. Fishov, M. Feingold, *Biophys. J.* **94**(1), 251 (2008). DOI 10.1529/biophysj.107.104398
- [34] H. Schlegel, C. Zaborosch, *General Microbiology*. Cambridge low price editions (Cambridge University Press, 1993). URL <https://books.google.co.in/books?id=DrHQtIbiunkC>
- [35] C.A. Athale, H. Chaudhari, *Bioinformatics* **27**(21), 2944 (2011). DOI 10.1093/bioinformatics/btr501
- [36] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, Y. Shao, *Science* **277**(5331), 1453 (1997)
- [37] N.A. Tanner, S.M. Hamdan, S. Jergic, K.V. Loscha, P.M. Schaeffer, N.E. Dixon, A.M. van Oijen, *Nature Struc. Mol. Biol.* **15**(2), 170 (2008). DOI 10.1038/nsmb.1381
- [38] B. Michel, H. Boubakri, Z. Baharoglu, M. LeMasson, R. Lestini, *DNA repair* **6**(7), 967 (2007). DOI 10.1016/j.dnarep.2007.02.016
- [39] M.M. Cox, M.F. Goodman, K.N. Kreuzer, D.J. Sherratt, S.J. Sandler, K.J. Mariani, *Nature* **404**(6773), 37 (2000)
- [40] S. Uphoff, R. Reyes-Lamothe, F. Garza de Leon, D.J. Sherratt, A.N. Kapandidis, *Proceedings of the National Academy of Sciences of the United States of America* **110**(20), 8063 (2013). DOI 10.1073/pnas.1301804110
- [41] H.E. Kubitschek, C.L. Woldringh, J. Bacteriol. **153**(3), 1379 (1983). URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=221788&tool=pmcentres>
- [42] E. Ben-Jacob, O. Schochet, A. Tenenbaum, I. Cohen, A. Czirok, T. Vicsek, *Nature* **368**(6466), 46 (1994)
- [43] T. Emonet, C.M. Macal, M.J. North, C.E. Wick-ersham, P. Cluzel, *Bioinformatics* **21**(11), 2714

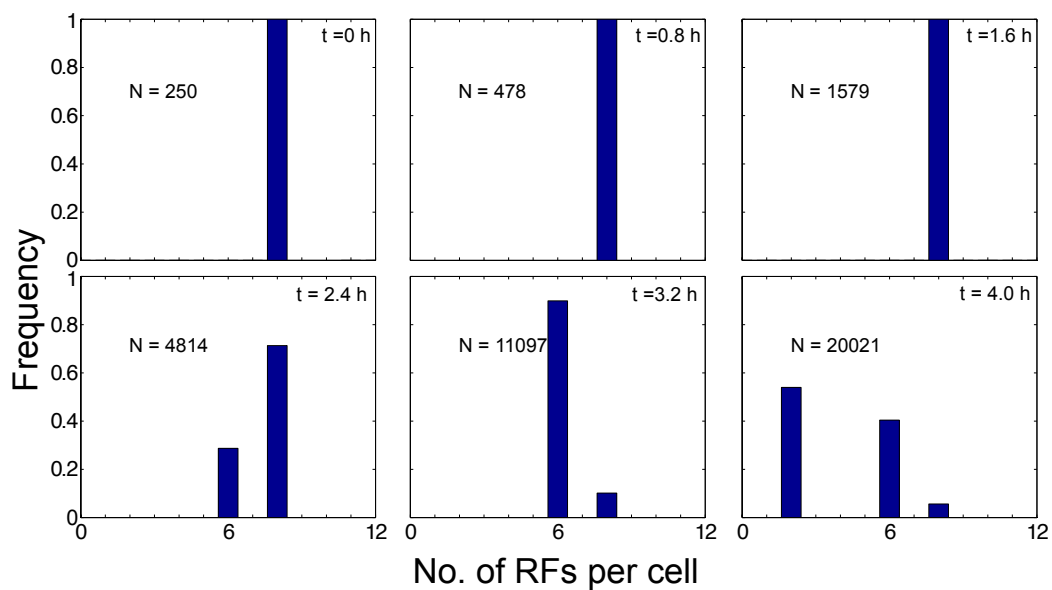
- (2005). DOI 10.1093/bioinformatics/bti391. URL <http://bioinformatics.oxfordjournals.org/content/21/11/2714.abstract>
- [44] T.E. Gorochowski, A. Matyjaszkiewicz, T. Todd, N. Oak, K. Kowalska, S. Reid, K.T. Tsaneva-Atanasova, N.J. Savery, C.S. Grierson, M. di Bernardo, PLoS ONE **7**(8), e42790 (2012). DOI 10.1371/journal.pone.0042790
 - [45] F.D. Farrell, O. Hallatschek, D. Marenduzzo, B. Waclaw, Phys. Rev. Lett. **111**(16), 168101 (2013)
 - [46] P. Mina, M. di Bernardo, N.J. Savery, K. Tsaneva-Atanasova, J R Soc Interface **10**(78), 20120612 (2013)



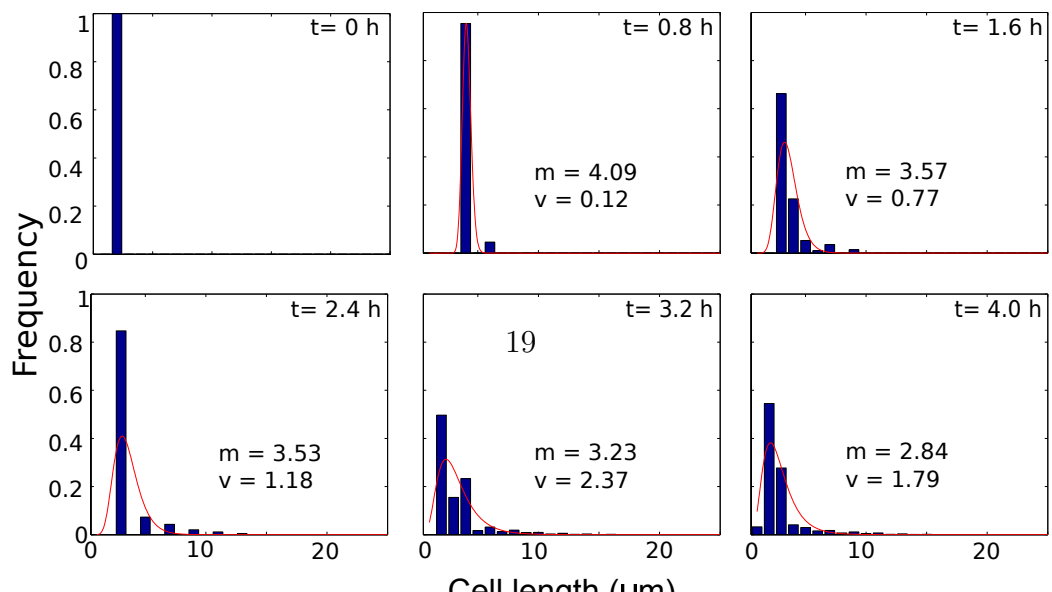
(a)



(b)



(c)



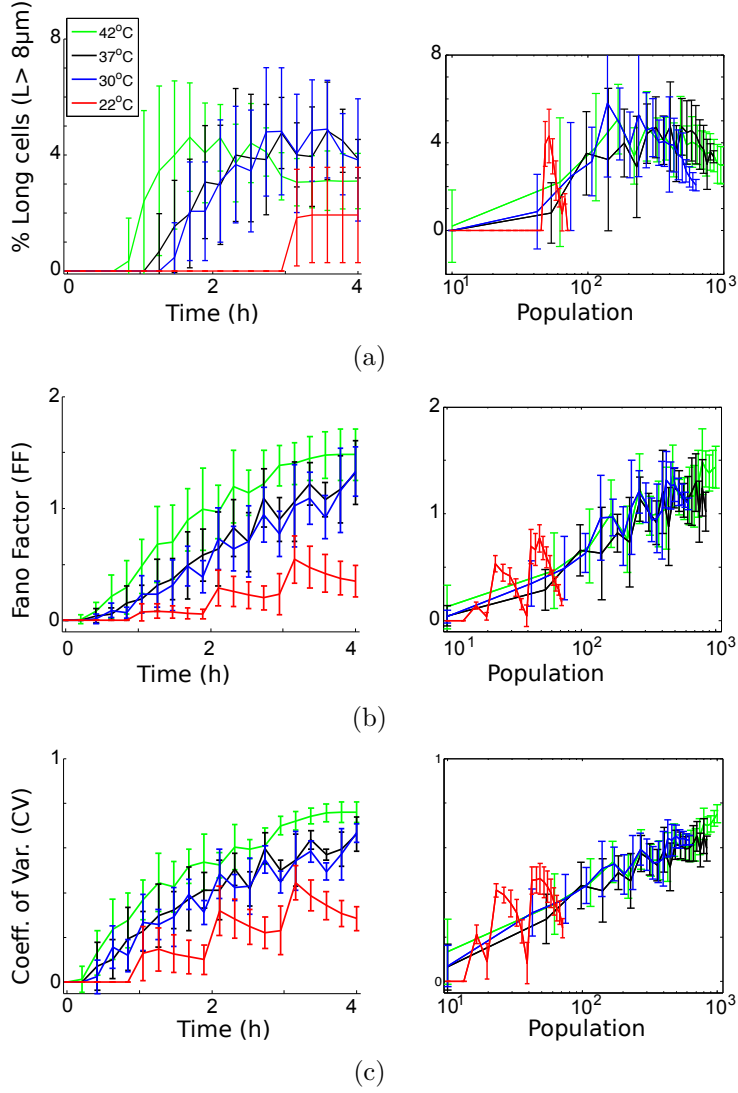


Figure 3: (a) Percentage long cells in the population, (b) fano factor and (c) coefficient of variation of cell lengths of the population are plotted for growth temperatures 22° (red), 30° (blue), 37° (black) and 42°C (green). The time (left) and population size (right) dependence of the measures are averages with errorbars representing s.d. from 25 iterations.

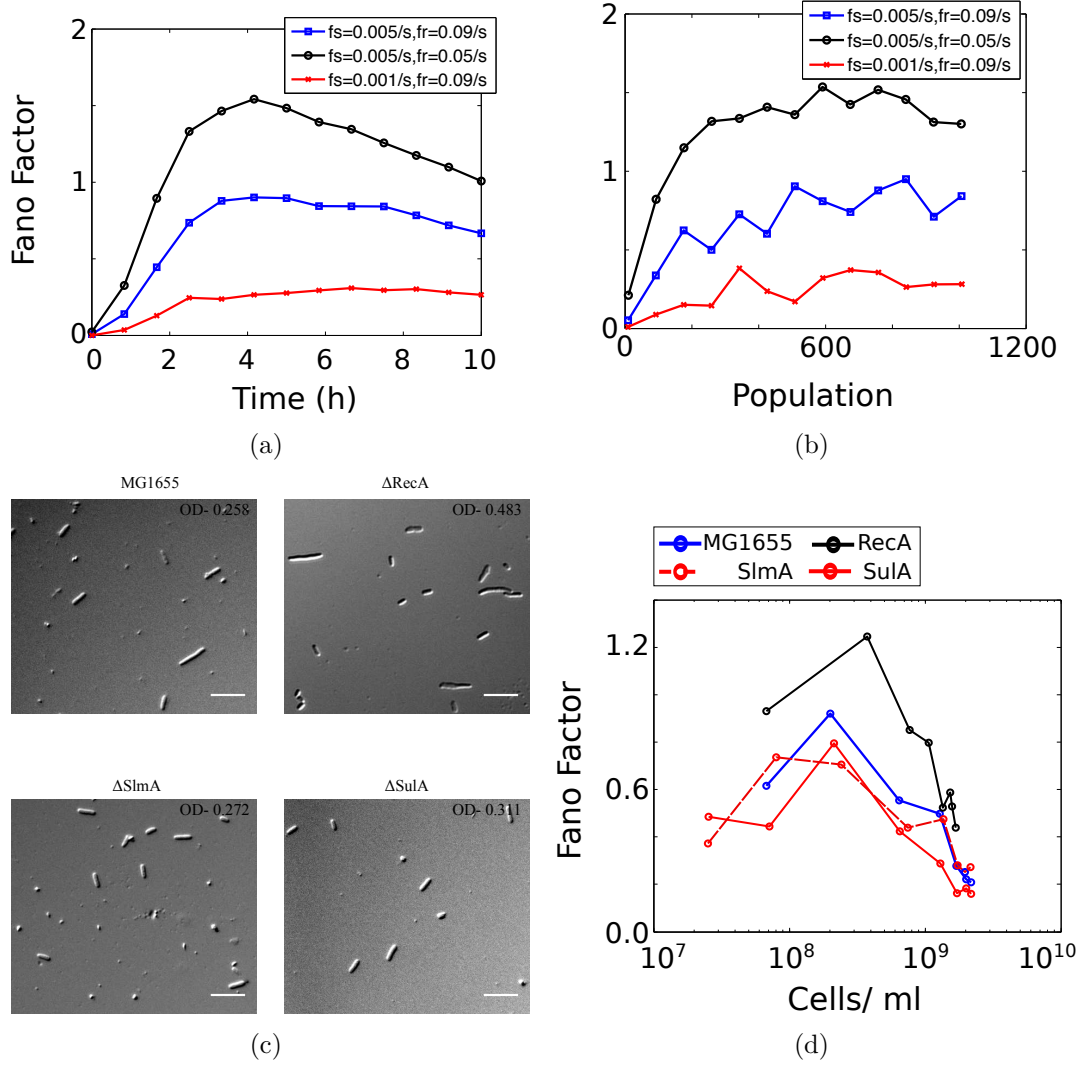


Figure 4: (a)-(b) The fano factor of simulated populations with $t_d^e = 20$ min (corresponding to 37°C) plotted as the fano factor of cell length as a function of time and population size. Conditions similar to wild-type ($f_s = 0.05$ s^{-1} and $f_r = 0.09$ s^{-1} , blue) are compared to a lower recovery frequency ($f_r = 0.05$, black) and lower stall frequency ($f_s = 0.001$, red). Representative microscopy images of wild-type (MG1655), ΔrecA , ΔslmA and ΔsulA from the same mid-log growth phase of cultures are presented (scale-bar 10 μm). The experimentally measured fano factor of cell length from a growth experiment are plotted as a function of cell density (cells/ml) for wild-type and the mutant strains.